

# Detecting Activations over Graphs using Spanning Tree Wavelet Bases

James Sharpnack <sup>\*1,3</sup>, Akshay Krishnamurthy <sup>†2</sup>, and Aarti Singh <sup>‡1</sup>

<sup>1</sup>Machine Learning Department, Carnegie Mellon University

<sup>2</sup>Computer Science Department, Carnegie Mellon University

<sup>3</sup>Statistics Department, Carnegie Mellon University

March 6, 2013

## Abstract

We consider the detection of activations over graphs under Gaussian noise, where signals are piece-wise constant over the graph. Despite the wide applicability of such a detection algorithm, there has been little success in the development of computationally feasible methods with prove-able theoretical guarantees for general graph topologies. We cast this as a hypothesis testing problem, and first provide a universal necessary condition for asymptotic distinguishability of the null and alternative hypotheses. We then introduce the spanning tree wavelet basis over graphs, a localized basis that reflects the topology of the graph, and prove that for any spanning tree, this approach can distinguish null from alternative in a low signal-to-noise regime. Lastly, we improve on this result and show that using the uniform spanning tree in the basis construction yields a randomized test with stronger theoretical guarantees that in many cases matches our necessary conditions. Specifically, we obtain near-optimal performance in edge transitive graphs,  $k$ -nearest neighbor graphs, and  $\epsilon$ -graphs.

## 1 Introduction

This paper focuses on the problem of detecting activations over a graph when observations are corrupted by noise. The problem of detecting graph-structured activations is relevant to many applications including identifying congestion in router and road networks, eliciting preferences in social networks, and detecting viruses in human and computer networks. Furthermore, these applications require that the method is scalable to large graphs. Luckily, computer science boasts a plethora of efficient graph based algorithms that we can adapt to the detection framework.

### 1.1 Contributions

In this paper, we will be testing if there is a non-zero piece-wise constant activation pattern on the graph given observations that are corrupted by Gaussian white noise. We show that correctly distinguishing the null and alternative hypotheses is impossible if the signal-to-noise ratio does not grow quickly with respect to the allowable number of discontinuities in the activation pattern (Section 2). Since a test based on the scan statistic which matches the observations with all possible activation patterns by brute force is infeasible, we propose a Haar wavelet basis construction for general graphs, which is formed by hierarchically dividing a spanning tree of the graph (Section 3). We find that the size and power of the test can be bounded in terms of the number of signal discontinuities and the spanning tree, immediately giving us a result for any spanning tree. We then propose choosing a spanning tree uniformly at random (this can be done efficiently), and show that this bound can be improved by a factor of the average effective resistance of the edges across which the signal is non-constant (Section 4). With this machinery in place we are able to show that for edge transitive graphs, such as lattices,  $k$ -nearest neighbor graphs, and  $\epsilon$  geometric random graphs, our test is nearly-optimal in that the upper bounds match the fundamental limits of detection up to logarithm factors (Section 5).

---

\*jsharpna@cs.cmu.edu

†akshaykr@cs.cmu.edu

‡aarti@cs.cmu.edu

## 1.2 Problem Setup

Consider an undirected graph  $G$  defined by a set of vertices  $V$  ( $|V| = n$ ) and undirected edges  $E$  ( $|E| = m$ ) which are unordered pairs of vertices. Throughout this study we will assume that the graph  $G$  is known. The statistical setting that we will address is the normal means model,

$$\mathbf{y} = \mathbf{x} + \epsilon$$

where  $\mathbf{x} \in \mathbb{R}^V$ ,  $\epsilon \sim N(0, \sigma^2 \mathbf{I}_V)$ , and  $\sigma^2$  is known. Specifically, we assume that there are parameters  $\rho, \mu$  (possibly dependent on  $n$ ) such that

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^V : |\{(v, w) \in E : x_v \neq x_w\}| \leq \rho, \|\mathbf{x}\| \geq \mu\}$$

defines the class of possible  $\mathbf{x}$ . Hence, the possible signals have few edges across which the values of  $\mathbf{x}$  differ. In graph-structured activation detection we are concerned with statistically testing the null and alternative hypotheses,

$$\begin{aligned} H_0 : \mathbf{y} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \\ H_1 : \mathbf{y} &\sim N(\mathbf{x}, \sigma^2 \mathbf{I}), \mathbf{x} \in \mathcal{X} \end{aligned} \tag{1}$$

$H_0$  represents business as usual while  $H_1$  encompasses all of the foreseeable anomalous activity. Let a test be a mapping  $T(\mathbf{y}) \in \{0, 1\}$ , where 1 indicates that we reject the null.

It is imperative that we control both the probability of false alarm, and the false acceptance of the null. To this end, we define our measure of risk to be

$$R(T) = \mathbb{E}_0[T] + \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbf{x}}[1 - T]$$

where  $\mathbb{E}_{\mathbf{x}}$  denote the expectation with respect to  $\mathbf{y} \sim N(\mathbf{x}, \sigma^2 \mathbf{I})$ . The test  $T$  may be randomized, in which case the risk is  $\mathbb{E}_T R(T)$ . Notice that if the distribution of the random test  $T$  is independent of  $\mathbf{x}$ , then  $\mathbb{E}_T \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbf{x}}[1 - T] = \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{T, \mathbf{x}}[1 - T]$ . This is the setting of Arias-Castro et al. [2011] which we should contrast to the Bayesian setup in Addario-Berry et al. [2010]. We will say that  $H_0$  and  $H_1$  are *asymptotically distinguished* by a test,  $T$ , if  $\lim_{n \rightarrow \infty} R(T) = 0$ . If such a test exists then  $H_0$  and  $H_1$  are asymptotically distinguished, otherwise they are asymptotically indistinguishable.

To aid us in our study we introduce some mathematical terminology. Let the edge-incidence matrix of  $G$  be  $\nabla \in \mathbb{R}^{E \times V}$  such that for  $(v, w) \in E$ ,  $\nabla_{(v, w), v} = 1$ ,  $\nabla_{(v, w), w} = -1$  (the order of  $(v, w)$  is chosen arbitrarily) and is 0 elsewhere. For a vector,  $\mathbf{w} \in \mathbb{R}^E$ ,  $\text{supp}(\mathbf{w}) = \{v \in V : \mathbf{w} \neq 0\}$  and  $\|\mathbf{w}\|_0 = |\text{supp}(\mathbf{w})|$ , so  $\|\nabla \mathbf{x}\|_0 \leq \rho$  for all  $\mathbf{x} \in \mathcal{X}$ . We will be constructing spanning trees  $\mathcal{T}$  of the graph  $G$ , which are connected subsets of  $E$  with no cycles. Furthermore, we will denote the edge-incidence matrix of  $\mathcal{T}$  as  $\nabla_{\mathcal{T}}$ .

## 1.3 Related Work

The statistical problem that we are addressing can be broadly classified as a high-dimensional Gaussian goodness-of-fit test. This is a well studied problem when the structure of  $H_1$  derives from a smooth function space such as an ellipsoid, Besov space or Sobolev space Ingster [1987], Ingster and Suslina [2003]. The function space  $\mathcal{X}$  that we are proposing is combinatorial in nature. This statistical problem has only recently been studied theoretically Addario-Berry et al. [2010], Arias-Castro et al. [2011], although to the best of our knowledge none have addressed the problem under arbitrary graph structure. More broadly, this work falls under the purview of multiple hypothesis testing, which has a rich history Benjamini and Hochberg [1995]. Unfortunately, aside from a few special cases Hall and Jin [2010], the multiple tests are assumed to be independent, making any such work not applicable to our setting.

In this paper, we evaluate our method by it's ability to distinguish  $H_0$  from  $H_1$ , however the procedure is based on constructing a wavelet basis over graphs which is relevant for other problems such as denoising and compression. Wavelets are multi-resolution bases that can represent inhomogeneous signals efficiently using a few non-zero wavelet coefficients which makes them attractive for denoising, compression and detection. As a result, they have been used extensively in mathematics, signal processing, statistics and physics Mallat [1999]. They have also been used with great success in statistics, with extensive theoretical guarantees Donoho and Johnstone [1995], Härdle et al. [1998], Vidakovic [1999]. Recently there has been some attention paid to developing wavelets for graphs. Unfortunately, most of these have either focused on graphs with a known hierarchical structure Gavish et al. [2010], Ram et al. [2011], Singh et al. [2010], or do not come with approximation or sparsifying properties that can be used for our class of graph functions  $\mathcal{X}$  Hammond et al. [2011], Coifman and Maggioni [2006].

## 2 Universal Lower Bound

In order to more completely understand the problem of detecting anomalous activity in graphs, we prove that there is a universal minimum signal strength under which  $H_0$  and  $H_1$  are asymptotically indistinguishable. The proof is based on a lemma developed in Arias-Castro et al. [2008], but the strategic use of this lemma is novel. Our construction of the ‘worst case’ prior gives a significantly tighter bound than would a more naive implementation. Indeed, it is interesting to note that the worst case prior is a uniform distribution of the largest unstructured signals that we are allowed in  $H_1$  that are nearly disjoint.

**Theorem 1.** *Let the maximum degree of  $G$  be  $d_{\max}$ . Consider the alternative,  $H_1$ , in which the cut size of each signal in  $\mathcal{X}$  is bounded by  $\rho$ , with  $\lim_{n \rightarrow \infty} \rho = \infty$  and  $\rho \leq \bar{d}n$ .  $H_0$  and  $H_1$  are asymptotically indistinguishable if*

$$\frac{\mu}{\sigma} = o\left(\sqrt{\min\left\{\frac{\rho}{d_{\max}}, \sqrt{n}\right\}}\right)$$

*Proof.* We begin by constructing a prior distribution over  $\mathcal{X}$ . This portion of the proof derives from the analysis in Arias-Castro et al. [2008] and closely mirrors that of Addario-Berry et al. [2010], Arias-Castro et al. [2011]. We will suppose that we have some subset  $\mathcal{S} \subseteq 2^V$  such that we will draw an  $S \in \mathcal{S}$  uniformly at random. Then the signal is constructed  $X = \frac{\mu}{\sqrt{|S|}} \mathbf{1}_S$  giving us a prior distribution  $\pi$  over  $\mathcal{X}$ . Call the Bayes risk  $R^*$ .

**Lemma 2.** *Addario-Berry et al. [2010] Let  $S$  and  $S'$  be drawn uniformly at random from  $\mathcal{S}$ . Then the Bayes risk  $R^*$  is bounded by*

$$R^* \geq 1 - \frac{1}{2} \sqrt{\mathbb{E} \exp\left(\frac{\mu^2}{2\sigma^2} \frac{|S \cap S'|}{\sqrt{|S||S'|}}\right)} - 1$$

Hence, if  $\mathbb{E} \exp\left(\frac{\mu^2}{2\sigma^2} \frac{|S \cap S'|}{\sqrt{|S||S'|}}\right) \rightarrow 1$ , then  $H_0$  and  $H_1$  are asymptotically indistinguishable. Let  $p = \lfloor \min\{\rho/d_{\max}, \sqrt{n}\} \rfloor$  and construct  $\mathcal{S}$  to be all subsets of  $V$  of size  $p$ . Then,

$$\mathbb{E} \exp \frac{\mu^2}{2\sigma^2} \frac{|S \cap S'|}{\sqrt{|S||S'|}} = \mathbb{E} \exp \frac{\mu^2}{2p\sigma^2} |S \cap S'|$$

Let  $\{z_i\}_{i=1}^p$  be Bernoulli trials with success probability  $p/n$ . We see that the distribution of  $|S \cap S'|$  is invariant under conditioning on  $S'$  and then it is equivalent to sampling without replacement from a population in which there are  $p$  successes. By Theorem 4 in Hoeffding [1963] we know that for  $t > 0$ ,  $\mathbb{E} e^{t|S \cap S'|} \leq \mathbb{E} e^{t \sum_{i=1}^p z_i}$ . Let  $t = \frac{\mu^2}{2p\sigma^2}$ , by the generating function of Bernoulli random variables,

$$\mathbb{E} e^{t \sum_{i=1}^p z_i} = \left(1 + \frac{p}{n} \left(e^{\frac{\mu^2}{2p\sigma^2}} - 1\right)\right)^p$$

By the assumption  $\frac{\mu^2}{\sigma^2} = o(p)$  so for any  $c > 0$  for  $n$  large enough

$$\left(1 + \frac{p}{n} \left(e^{\frac{\mu^2}{2p\sigma^2}} - 1\right)\right)^p \leq \left(1 + c \frac{p}{n}\right)^p \leq \left(1 + \frac{c}{p}\right)^p \rightarrow e^c$$

because  $p \leq \sqrt{n}$ . Hence,  $\mathbb{E} e^{t|S \cap S'|} \rightarrow 1$ . All that remains is to notice that the cut sizes of  $S \in \mathcal{S}$  are bounded by  $\rho$  because the cut sizes are bounded by  $pd_{\max} \leq \rho$ .  $\square$

## 3 Spanning Tree Wavelets

In this section, we present an algorithm for constructing a wavelet basis given a spanning tree and we characterize its performance for the detection problem (1).

Informally, we would like to construct a basis  $\mathbf{B}$  for which each edge  $e \in \mathcal{T}$  is activated by very few basis elements, where we say that an edge  $e$  is activated by element  $\mathbf{b}$  if  $e \in \text{supp}(\nabla_{\mathcal{T}} \mathbf{b})$ . As we will show, upper bounding

the number of basis elements that activate any edge will be essential in analyzing the performance of our estimator  $\|\mathbf{By}\|_\infty$ .

We construct our wavelet basis  $\mathbf{B}$  recursively, by first finding a seed vertex in the spanning tree such that the subtrees adjacent to the seed have at most  $\lceil n/2 \rceil$  vertices and then by including basis elements localized on these subtrees in  $\mathbf{B}$ . We recurse on each subtree, adding higher-resolution elements to our basis, and consequently constructing a complete wavelet basis. The first phase of the algorithm ensures that the depth of the recursion is at most  $\lceil \log n \rceil$  and the second ensures that each edge is activated by at most  $\lceil \log d \rceil$  basis elements per recursive call. Combining these two shows that each edge is activated by at most  $\lceil \log d \rceil \lceil \log n \rceil$  basis elements.

Finding a balancing vertex in the tree parallels the technique in Pearl and Tarsi [1986], which finds a balancing edge. The algorithm starts from any vertex  $v \in \mathcal{T}$  and moves along  $\mathcal{T}$  to a neighboring vertex  $w$  that lies in the largest connected component of  $\mathcal{T} \setminus v$ . The algorithm repeats this process (moving from  $v$  to  $w$ ) until the largest connected component of  $\mathcal{T} \setminus w$  is larger than the largest connected component of  $\mathcal{T} \setminus v$  at which point it returns  $v$ . We call this the *FindBalance* algorithm.

Once we have a balancing vertex  $v$ , we form wavelets that are constant over the connected components of  $\mathcal{T} \setminus v$  such that any vertex is supported by at most  $\log d$  wavelets. Let  $d_v$  be the degree of the balancing vertex  $v$  and let  $c_1, \dots, c_{d_v}$  be the connected components of  $\mathcal{T} \setminus v$  (with  $v$  added to the smallest component). Our algorithm acts as if  $c_1, \dots, c_{d_v}$  form a chain structure and constructs the Haar wavelet basis over them. We call this algorithm *FormWavelets*:

1. Let  $C_1 = \cup_{i \leq d_v/2} c_i$  and  $C_2 = \cup_{i > d_v/2} c_i$
2. Form the following basis element and add it to  $\mathbf{B}$ :

$$\mathbf{b} = \frac{\sqrt{|C_1||C_2|}}{\sqrt{|C_1|+|C_2|}} \left[ \frac{1}{|C_1|} \mathbf{1}_{C_1} - \frac{1}{|C_2|} \mathbf{1}_{C_2} \right]$$

3. Recurse at (1) with the subcomponents of  $C_1$  and  $C_2$  with partitions  $\{c_i\}_{i \leq p/2}$  and  $\{c_i\}_{i > p/2}$  respectively.

Our algorithm recursively constructs basis elements using the *FindBalance* and *FormWavelets* routines on subtrees of  $\mathcal{T}$ . We initialize  $\mathcal{T}$  to be a spanning tree of the graph and start with no elements in our basis.

1. Let  $v$  be the output of *FindBalance* applied to  $\mathcal{T}$ .
2. Let  $\mathcal{T}_1, \dots, \mathcal{T}_{d_v}$  be the connected components of  $\mathcal{T} \setminus v$  and add  $v$  to the smallest component.
3. Add the basis elements constructed in *FormWavelets* when applied to  $\mathcal{T}_1, \dots, \mathcal{T}_{d_v}$
4. For each  $i \in [d_v]$ , recursively apply (1) - (4) on  $\mathcal{T}_i$  as long as  $|\mathcal{T}_i| > 2$ .

As we will see, controlling the sparsity,  $\|\mathbf{Bx}\|_0$  is essential in analyzing the performance of the estimator  $\|\mathbf{By}\|_\infty$ . The main theoretical guarantee of our basis construction algorithm is that signals with small cuts in  $G$  are sparse in  $\mathbf{B}$ . Specifically, we prove the following key lemma in the appendix:

**Lemma 3.** *Let  $\nabla$  be the incidence matrix of  $G$  and  $\nabla_{\mathcal{T}}$  be the incidence matrix of  $\mathcal{T}$  (where  $\mathcal{T}$  has degree at most  $d$ ). Then  $\|\nabla \mathbf{x}\|_0$  is the cut size of pattern  $\mathbf{x} \in \mathbb{R}^{V(G)}$ . Then for any  $\mathbf{x} \in \mathbb{R}^{V(G)}$ ,*

$$\|\mathbf{Bx}\|_0 \leq \|\nabla_{\mathcal{T}} \mathbf{x}\|_0 \lceil \log d \rceil \lceil \log n \rceil \leq \|\nabla \mathbf{x}\|_0 \lceil \log d \rceil \lceil \log n \rceil \quad (2)$$

Equipped with Lemma 3 we can now characterize the performance of the estimator  $\|\mathbf{By}\|_\infty$  on any signal  $\mathbf{x}$ . Our bound depends on the choice of spanning tree  $\mathcal{T}$ , specifically via the quantity  $\|\nabla_{\mathcal{T}} \mathbf{x}\|_0$ , the cut size of  $\mathbf{x}$  in  $\mathcal{T}$ . The proof of the following can be found in the appendix.

**Theorem 4.** *Perform the test in which we reject the null if  $\|\mathbf{By}\|_\infty > \tau$ . Set  $\tau = \sigma \sqrt{2 \log(n/\delta)}$ . If*

$$\frac{\mu}{\sigma} \geq \sqrt{2 \|\nabla_{\mathcal{T}} \mathbf{x}\|_0 \lceil \log d \rceil \lceil \log n \rceil} (\sqrt{\log(1/\delta)} + \sqrt{\log(n/\delta)}) \quad (3)$$

*then under  $H_0$ ,  $\mathbb{P}\{\text{Reject}\} \leq \delta$ , and under  $H_1$ ,  $\mathbb{P}\{\text{Reject}\} \geq 1 - \delta$ .*

**Remark 5.** *For any tree we have  $\|\nabla_{\mathcal{T}} \mathbf{x}\|_0 \leq \|\nabla \mathbf{x}\|_0$  for all patterns  $\mathbf{x}$ , so that for the sparse cut alternative we can have both Type I and Type II errors  $\leq \delta$  as long as:*

$$\frac{\mu}{\sigma} \geq \sqrt{2 \rho \lceil \log d \rceil \lceil \log n \rceil} (\sqrt{\log(1/\delta)} + \sqrt{\log(n/\delta)}) \quad (4)$$

## 4 Uniform Spanning Tree Basis

The uniform spanning tree (UST) is a spanning tree generation technique that we will use to construct wavelet bases. We will first examine the deep connection between electrical networks, USTs and random walks. Because the UST is randomly generated, the test statistic,  $\|\mathbf{B}_{\mathcal{T}}\mathbf{y}\|$  when conditioned on  $\mathbf{y}$  will also be random. Due to results from cut sparsification, we can relate the performance of the UST wavelet detector to effective resistances.

### 4.1 Cuts and Effective Resistance

Effective resistances have been extensively studied in electrical network theory. We define the combinatorial Laplacian of  $G$  to be  $\Delta = \nabla^\top \nabla$ . A *potential difference* is any  $\mathbf{z} \in \mathbb{R}^E$  such that it satisfies *Kirchoff's potential law*: the total potential difference around any cycle is 0. Algebraically, this means that  $\exists \mathbf{x} \in \mathbb{R}^V$  such that  $\nabla \mathbf{x} = \mathbf{z}$ . The *Dirichlet Principle* states that any solution to the following program gives an absolute potential  $\mathbf{x}$  that satisfies Kirchoff's potential law:

$$\min. \mathbf{x}^\top \Delta \mathbf{x} \text{ s.t. } \mathbf{x}_S = \mathbf{v}_S$$

for source/sinks  $S \subset V$  and some voltage constraints  $\mathbf{v}_S \in \mathbb{R}^S$ . The realized objective  $\mathbf{x}^\top \Delta \mathbf{x}$  is known as the *total energy* of the system. By Lagrangian calculus, the solution to the above program is given by  $\mathbf{x} = \Delta^\dagger \mathbf{v}$  where  $\dagger$  indicates the Moore-Penrose pseudoinverse. The effective resistance is the total energy of a system in which  $v, w \in V$  are the source and sink respectively and a unit flow from  $v$  to  $w$  is induced. Hence, the effective resistance between  $v$  and  $w$  is  $r_{v,w} = (\delta_v - \delta_w)^\top \Delta^\dagger (\delta_v - \delta_w)$ , where  $\delta_v$  is the Dirac delta function.

A massively useful characterization of effective resistance is the random walk interpretation. Let  $X_t$  be the location of a random walker on  $G$  at time  $t$ . The hitting time  $H(v, w)$  is then

$$H(v, w) = \mathbb{E}[\min\{t > 0 : X_t = w\} | X_0 = v]$$

We find that the effective resistance is related to the hitting time by,

$$r_{v,w} = \frac{H(v, w) + H(w, v)}{2m}$$

The numerator is also known as the commute time. As we will see, this characterization of effective resistance is useful when bounding it for specific graph models.

### 4.2 UST Wavelet Detector

In our framework, we are given the opportunity to evaluate our test according to our random algorithm. We will now examine the performance of the spanning tree wavelet detector, when the spanning tree is drawn according to a UST. First, we will explore the construction of the UST and examine key properties. The UST is a random spanning tree, chosen uniformly at random from the set of all distinct spanning trees. The foundational Matrix-Tree theorem Kirchhoff [1847] describes the probability of an edge being included in the UST. The following lemma can be found in Lovász [1993] and Lyons and Peres [2000].

**Lemma 6.** *Let  $G$  be a graph and  $\mathcal{T}$  a draw from  $UST(G)$ .*

$$\mathbb{P}\{e \in \mathcal{T}\} = r_e$$

Hence, we can expect that for a given cut in the graph, that the cut size in the tree will look like the sum of edge effective resistances. While it is infeasible to enumerate all spanning trees of a graph, the Aldous-Broder algorithm is an efficient method for generating a draw from  $UST(G)$  Aldous [1990]. The algorithm simulates a random walk on  $G$ ,  $\{X_t\}$ , stops when all of the vertices have been visited, and defines the spanning tree  $\mathcal{T}$  by the edges  $\{(X_{H(X_0, v)-1}, v) : v \in V\}$ .

In order to control  $\|\nabla_{\mathcal{T}} \mathbf{x}\|_0$ , we need to control the overlap between a cut and the UST. Clearly the UST does not independently sample edges, but it does have the well documented property of negative association, that the inclusion of an edge decreases the probability that another edge is included. The following lemma states a concentration result for the UST, based on negative association, and can be found in Fung and Harvey [2010]. The proof is a simple extension of the concentration results in Gandhi et al. [2006].

**Lemma 7.** Let  $B \subset E$  be a fixed subset of edges, and  $|\mathcal{T} \cap B|$  denote the number of edges in  $\mathcal{T}$  also in  $B$ .

$$\mathbb{P}\{|\mathcal{T} \cap B| \geq (1 + \delta) \sum_{e \in B} r_e\} \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\sum_{e \in B} r_e}$$

We use this result to give conditions under which the UST wavelet detector asymptotically distinguishes  $H_0$  from  $H_1$ .

**Theorem 8.** Let  $r_{\max} = \max_{\mathbf{x} \in \mathcal{X}} \sum_{e \in \text{supp}(\nabla \mathbf{x})} r_e$  (the maximum effective resistance of a cut in  $\mathcal{X}$ ). If

$$\frac{\mu}{\sigma} = \omega \left( \sqrt{r_{\max} \log d \log n} \right)$$

then  $H_0$  and  $H_1$  are asymptotically distinguished by the test statistic  $\|\mathbf{B}\mathbf{y}\|_\infty$  where  $\mathbf{B}$  is the UST wavelet basis.

*Proof.* Let  $r_B = \sum_{e \in B} r_e$  for  $B \subset E$ . By some basic calculus, and the fact that  $\log(1 + x) \geq x/(1 + x/2)$ , we see that

$$\left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{r_B} \leq \exp \left( -\frac{\delta^2 r_B}{2 + \delta} \right)$$

Rewriting the Lemma 7, we obtain with probability  $> 1 - \gamma$

$$|\mathcal{T} \cap B| \leq r_B + \sqrt{2r_B \log \frac{1}{\gamma} + \frac{1}{4}(\log \frac{1}{\gamma})^2} + \frac{1}{2} \log \frac{1}{\gamma} \leq \left( r_B + \sqrt{2r_B \log \frac{1}{\gamma} + \log \frac{1}{\gamma}} \right)$$

Now, because  $\|\nabla_{\mathcal{T}} \mathbf{x}\|_0 = |\mathcal{T} \cap B|$  for  $B = \text{supp}(\nabla_{\mathcal{T}} \mathbf{x})$ , we know by Theorem 4 if

$$\frac{\mu}{\sigma} = \omega \left( \sqrt{\left( r_B + \sqrt{2r_B \log \frac{1}{\gamma} + \log \frac{1}{\gamma}} \right) \log d \log n} \right)$$

then  $H_0$  and  $H_1$  are asymptotically distinguished and the result follows because we guarantee this uniformly for all such  $B$ .  $\square$

## 5 Specific Graph Models

In this section we study our detection problem for several different families of graphs. Specifically, we control the effective resistance  $r_e$  for each graph family, which when combined with Theorem 8 gives a lower bound on the SNR for which  $\|\mathbf{B}\mathbf{y}\|_\infty$  asymptotically distinguishes  $H_0$  and  $H_1$ .

In Theorem 8, we showed that the consistency regime depends on the effective resistances of the cuts induced by the class of signals  $\mathcal{X}$ . On its own, it is not immediately clear that this result is an improvement over the bound in Remark 5 that we would obtain from any spanning tree. However, Foster's theorem highlights why we expect the effective resistance to be less than the cut size.

**Theorem 9** (Foster's Theorem Foster [1949], Tetali [1991]).

$$\sum_{e \in E(G)} r_e = n - 1$$

Hence, if we select an edge uniformly at random from the graph, we expect its effective resistance to be  $(n - 1)/m \approx \bar{d}^{-1}$  (the reciprocal of the average degree) where  $m \triangleq |E(g)|$ . Indeed, in several example graphs we can formalize this intuition and give an improvement over Remark 5.

We complement these results with two types of simulations verifying different aspects of our theory. The first verifies the upper bound in Lemma 3 for a variety of graph models by plotting  $\|\mathbf{B}\mathbf{x}\|_0$  versus  $\rho \log(d) \log(n)$  for several randomly generated signals. These plots (see Figure 1) demonstrate the validity of our bound since in all cases  $\|\mathbf{B}\mathbf{x}\|_0 \leq \rho \log(d) \log(n)$ , but, more importantly, the readily-observable linear relationship between these two quantities suggests that one should not expect an improvement on this bound by more than a constant factor.

The second simulation verifies the performance of our spanning tree wavelets detector on various graph models. In Figure 2, we plot the power of our test statistic (with Type I error fixed at 5%) as a function of signal strength  $\mu$  for several values of  $n$ , where we allow  $\rho$  to scale with  $n$  to ensure a non-empty  $\mathcal{X}$ . These simulations demonstrate that as expected for sufficiently large signal strength, our statistic can separate  $H_0$  from  $H_1$ . More importantly, the threshold signal strength for which detection is possible increases with  $n$  and  $\rho$ , as predicted by our theory.



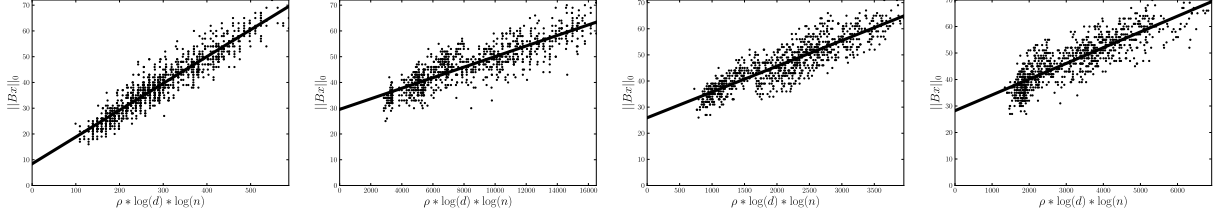


Figure 1: Spanning tree wavelet basis sparsity as a function of  $\rho \log d \log n$  for, from left to right, 2-dimensional torus, complete,  $k$ -NN, and  $\epsilon$  graphs. Linear fits have slopes: 0.10, 0.0021, 0.010, 0.0059 and  $R^2$  coefficients: 0.88 0.72, 0.76, 0.71 respectively.

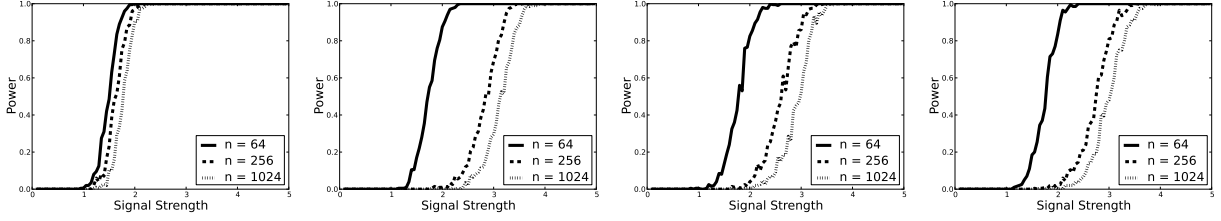


Figure 2: Power as a function of signal strength for different values of  $n$  for 2-dimensional torus, complete,  $k$ -NN, and  $\epsilon$  graphs.  $\rho$  scales like  $\sqrt{n}$ ,  $n$ ,  $n^{2/3}$  and  $n^{4/5}$  respectively.

## 5.1 Edge Transitive Graphs

An edge transitive graph,  $G$ , is one such that for any edges  $e_0, e_1$ , there is a graph automorphism that maps  $e_0$  to  $e_1$ . Examples of edge transitive graphs include the  $l$ -dimensional torus and the complete graph  $K_n$ . For such a graph, every edge has the same effective resistance, and Foster's Theorem then shows that  $r_e = (n-1)/m$  where  $m$  is the number of edges. Moreover since edge transitive graphs must be  $d$ -regular for some degree  $d$ , we see that  $m = \Theta(nd)$  so the  $r_e = \Theta(1/d)$ . This leads us to the following corollary, which we note matches the lower bound in Theorem 1 modulo logarithmic terms if  $\rho/d \leq \sqrt{n}$ :

**Corollary 10.** *Let  $G$  be edge transitive with common degree  $d$ . Then for each edge  $e \in E(G)$ ,  $r_e = (n-1)/m$ . Consider the hypothesis testing problem (1) where the set  $\mathcal{X}$  is parameterized by  $\rho$ . If:*

$$\frac{\mu}{\sigma} = \omega \left( \sqrt{\frac{\rho}{d} \log d \log n} \right)$$

*Then the UST wavelet detector,  $\|\mathbf{By}\|_{\infty}$ , asymptotically distinguishes  $H_0$  and  $H_1$ .*

## 5.2 kNN Graphs

Oftentimes in applications, the graph topology is derived from data. In this case, the randomness of the data means that the graph itself is inherently random. Commonly, these graphs are modeled as random geometric graphs, and in this section we will devote our attention to the *symmetric  $k$ -nearest neighbor graphs*. Specifically, suppose that  $\mathbf{z}_1, \dots, \mathbf{z}_n$  are drawn i.i.d. from a density  $p$  supported over  $\mathbb{R}^d$ . Then we form the graph  $G$  over  $[n]$  by connecting vertices  $i, j$  if  $\mathbf{z}_i$  is amongst the  $k$ -nearest neighbors of  $\mathbf{z}_j$  or vice versa. Some regularity conditions of  $p$  are needed for our results to hold; they can be found in Von Luxburg et al. [2010].

To bound the effective resistance  $r_e$ , Corollary 9 in Von Luxburg et al. [2010] shows that  $H_{ij}/2m \rightarrow 1/d_j$  and by the definition of  $r_e$  we see that  $r_{ij} \rightarrow \frac{1}{d_i} + \frac{1}{d_j} \leq \frac{2}{k}$ , since  $d_i \geq k$  for each  $i$ . A formal analysis leads to the following corollary, which we prove in Appendix B with more precise concentration arguments:

**Corollary 11.** *Let  $G$  be a  $k$ -NN graph with  $k/n \rightarrow 0$  and  $k(k/n)^{2/d} \rightarrow \infty$  and where the density  $p$  satisfies the regularity conditions in Von Luxburg et al. [2010]. Consider the hypothesis testing problem (1) where the set  $\mathcal{X}$  is*

parameterized by  $\rho$ . If:

$$\frac{\mu}{\sigma} = \omega(\sqrt{\rho/k \log d \log n})$$

Then the UST wavelet detector,  $\|\mathbf{By}\|_\infty$ , asymptotically distinguishes  $H_0$  and  $H_1$ .

### 5.3 $\epsilon$ -Graphs

The  $\epsilon$ -graph is another widely used random geometric graph in machine learning and statistics. As with the  $k$ -NN graph, the vertices are embedded into  $\mathbb{R}^d$  and edges are added between pairs of vertices that are within distance  $\epsilon$  of each other. As with the  $k$ -NN graph, Corollary 8 from Von Luxburg et al. [2010] shows that  $H_{ij} \rightarrow m/d_j$  for each pair of vertices. This leads us to believe that  $r_{ij} \rightarrow 1/(d_i) + 1/(d_j)$ . If the density  $p$  from which we draw data points is bounded from below by some constant, then we can uniformly lower bound all of the degrees  $d_i$  using fairly elementary concentration results, which results in an upper bound on  $r_e$ . Formalizing this intuition, we have the following corollary, which we prove in Appendix B:

**Corollary 12.** *Let  $G$  be a  $\epsilon$ -graph with points  $X_1, \dots, X_n$  drawn from a density  $p$ , which satisfies the regularity conditions in Von Luxburg et al. [2010] and is lower bounded by some constant  $p_{\min}$  (independent of  $n$ ). Let  $\epsilon \rightarrow 0$ ,  $n\epsilon^{d+2} \rightarrow \infty$  and consider the hypothesis testing problem (1) where the set  $\mathcal{X}$  is parameterized by  $\rho$ . If:*

$$\frac{\mu}{\sigma} = \omega(\sqrt{\frac{\rho}{n\epsilon^d} \log d \log n})$$

Then  $\|\mathbf{By}\|_\infty$  asymptotically distinguishes  $H_0$  and  $H_1$ .

## 6 Discussion

We studied the detection of piece-wise constant activation patterns over graphs, and provided a necessary condition for the asymptotic distinguishability of signals that are assumed to have few discontinuities. We gave a novel spanning tree wavelet construction, that is the extension of the Haar wavelet basis, for arbitrary graphs. While it achieves strong theoretical performance for detection, the spanning tree wavelet construction could be of independent interest for compression and denoising. The uniform spanning tree wavelet detector was shown to have strong theoretical guarantees that in many cases gives us near optimal performance. This means that under adversarial choice of signal, our randomized algorithm asymptotically distinguishes  $H_0$  from  $H_1$ . Alternatively, this means that for any given signal (non-adversarial setting) that the vast majority of spanning trees induce detectors that asymptotically distinguish  $H_0$  from  $H_1$  for low signal-to-noise ratios.

## Acknowledgements

This research is supported in part by AFOSR under grant FA9550-10-1-0382 and NSF under grant IIS-1116458. AK is supported in part by an NSF Graduate Research Fellowship.

## References

- L. Addario-Berry, N. Broutin, L. Devroye, and G. Lugosi. On combinatorial testing problems. *The Annals of Statistics*, 38(5):3063–3092, 2010.
- D. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM Journal on Discrete Mathematics*, 3(4):450–465, 1990.
- E. Arias-Castro, E.J. Candes, H. Helgason, and O. Zeitouni. Searching for a trail of evidence in a maze. *The Annals of Statistics*, 36(4):1726–1757, 2008.
- E. Arias-Castro, E.J. Candes, and A. Durand. Detection of an anomalous cluster in a network. *The Annals of Statistics*, 39(1):278–304, 2011.



- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.
- R Coifman and M Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21(1):53–94, 2006. URL <http://linkinghub.elsevier.com/retrieve/pii/S106352030600056X>.
- D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, pages 1200–1224, 1995.
- R.M. Foster. The average impedance of an electrical network. *Contributions to Applied Mechanics (Reissner Anniversary Volume)*, pages 333–340, 1949.
- W.S. Fung and N.J.A. Harvey. Graph sparsification by edge-connectivity and random spanning trees. *Arxiv preprint arXiv:1005.0265*, 2010.
- R. Gandhi, S. Khuller, S. Parthasarathy, and A. Srinivasan. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM (JACM)*, 53(3):324–360, 2006.
- Matan Gavish, Boaz Nadler, and Ronald R. Coifman. Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning. In *ICML*, pages 367–374, 2010.
- P. Hall and J. Jin. Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics*, 38(3):1686–1732, 2010.
- D.K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.
- W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*. Springer Berlin, 1998.
- W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, pages 13–30, 1963.
- Y.I. Ingster. Minimax testing of nonparametric hypotheses on a distribution density in the  $l_p$  metrics. *Theory of Probability and its Applications*, 31:333, 1987.
- Y.I. Ingster and I.A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Verlag, 2003.
- G. Kirchhoff. Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird. *Annalen der Physik*, 148(12):497–508, 1847.
- L. Lovász. Random walks on graphs: A survey. *Combinatorics, Paul Erdos is Eighty*, 2(1):1–46, 1993.
- R. Lyons and Y. Peres. Probability on trees and networks. Book in preparation., 2000.
- S.G. Mallat. *A wavelet tour of signal processing*. Academic Pr, 1999.
- J. Pearl and M. Tarsi. Structuring causal trees. *Journal of Complexity*, 2(1):60–77, 1986.
- I. Ram, M. Elad, and I. Cohen. Generalized tree-based wavelet transform. *Signal Processing, IEEE Transactions on*, 59(9):4199–4209, 2011.
- A. Singh, R.D. Nowak, and R. Calderbank. Detecting weak but hierarchically-structured patterns in networks. *Arxiv preprint arXiv:1003.0205*, 2010.
- P. Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(1):101–109, 1991.
- B. Vidakovic. *Statistical modeling by wavelets*. Wiley Online Library, 1999.
- Ulrike Von Luxburg, Agnes Radl, and Matthias Hein. Hitting and commute times in large graphs are often misleading. *ReCALL*, 2010. URL <http://arxiv.org/abs/1003.1266>.

## A Proofs for Section 3

### A.1 Proof of Lemma 3

Before we proceed with the proof, we state and prove two results on the performance of the algorithm:

**Lemma 13.** *Let  $\mathcal{T}$  be a tree. FindBalance returns a vertex  $v$  such that the largest connected component of  $\mathcal{T} \setminus v$  is of size at most  $\lceil |\mathcal{T}|/2 \rceil$  in  $O(|\mathcal{T}|)$  time.*

*Proof.* Let the objective be the size of the largest connected components of  $\mathcal{T} \setminus v$ . Every move in *FindBalance* reduces the objective by at least 1 and the objective can be at most  $|\mathcal{T}| - 1$  so it must terminate in less than  $|\mathcal{T}|$  moves. Now at any step of *FindBalance*, if the objective is greater than  $\lceil |\mathcal{T}|/2 \rceil$ , the cumulative size of the remaining connected components is less than  $\lfloor |\mathcal{T}|/2 \rfloor$ . Hence, in the next step the connected component formed by these is less than  $\lceil |\mathcal{T}|/2 \rceil$ . Thus, the program cannot terminate at a move directly after the objective is greater than  $\lceil |\mathcal{T}|/2 \rceil$ .  $\square$

We will also require the following claim. Indeed, controlling the depth of the recursion in the wavelet construction is the sine qua non for controlling the sparsity,  $\|\mathbf{B}\mathbf{x}\|_0$ .

**Claim 14.** *The wavelet construction has recursion depth at most  $\lceil \log d \rceil \lceil \log n \rceil$ .*

*Proof.* Whenever *FormWavelet* is applied it increases the number of activated height of the dendrogram by at most  $\lceil \log d \rceil$ . By lemma 13 the size of the remaining components is halved, so the algorithm terminates in at most  $\lceil \log n \rceil$  steps.  $\square$

*Proof of Lemma 3.* We will show that any edge  $e \in \mathcal{T}$  is activated by at most  $\lceil \log d \rceil \lceil \log n \rceil$  basis elements in  $\mathbf{B}$ , and this will imply the result. We will say that an edge  $e$  is activated by a basis element  $\mathbf{b}$  if  $e \subseteq \text{supp}(\nabla_{\mathcal{T}} \mathbf{b})$ . It follows that for a basis element  $\mathbf{b}$ , if  $\mathbf{b}^T \mathbf{x} \neq 0$  then  $\exists e$  that is activated by  $\mathbf{b}$ . Let  $\text{activations}(e)$  be the number of basis elements that activates  $e$  ( $\text{activations}(e) = 0$  if  $e \notin \text{supp}(\nabla_{\mathcal{T}} \mathbf{x})$ ). We then have

$$\|\mathbf{B}\mathbf{x}\|_0 \leq \sum_{e \in \text{supp}(\nabla_{\mathcal{T}} \mathbf{x})} \text{activations}(e)$$

Consider some edge  $e$ . If  $e$  is activated by some subtree  $\mathcal{T}_{\text{sub}}$  (we use this interchangeably with being activated by the basis element formed by partitioning  $\mathcal{T}_{\text{sub}}$  into two groups), then it can be activated by at most one of  $\mathcal{T}_{\text{sub}}$ 's subtrees. This implies that  $\text{activations}(e)$  is upper bounded by the depth of the recursion. By the claim, we find that,

$$\|\mathbf{B}\mathbf{x}\|_0 \leq \sum_{e \in \text{supp}(\nabla_{\mathcal{T}} \mathbf{x})} \lceil \log d \rceil \lceil \log n \rceil \leq \|\nabla_{\mathcal{T}} \mathbf{x}\|_0 \lceil \log d \rceil \lceil \log n \rceil$$

Proving the first claim. The second claim is obvious from the fact that  $\mathcal{T}$  contains a subset of the edges in  $\mathcal{G}$ , so every cut has larger cut size in  $\mathcal{G}$  than it does in  $\mathcal{T}$ .  $\square$

### A.2 Proof of Theorem 4

*Proof.* Under the null  $\mathbf{x} = 0$ , and we have that

$$\|\mathbf{B}\mathbf{y}\|_{\infty} = \|\mathbf{B}\boldsymbol{\epsilon}\|_{\infty} < \sigma \sqrt{2 \log(n/\delta)}$$

with probability at least  $1 - \delta$ . So, as long as  $\tau = \sigma \sqrt{2 \log(n/\delta)}$  then we control the probability of false alarm (type 1 error). For a element  $\mathbf{x}$  of the alternative, let the index,  $i^*$ , achieve the maximum of  $\mathbf{B}\mathbf{x}$  (i.e.  $\|\mathbf{B}\mathbf{x}\|_{\infty} = |\mathbf{B}\mathbf{x}|_{i^*}$ ). Then  $|\mathbf{B}\mathbf{y}|_{i^*} \geq |\mathbf{B}\mathbf{x}|_{i^*} - \sigma \sqrt{2 \log(1/\delta)}$  with probability at least  $1 - \delta$  and

$$|\mathbf{B}\mathbf{x}|_{i^*}^2 = \|\mathbf{B}\mathbf{x}\|_{\infty}^2 \geq \frac{\sum_{i: (\mathbf{B}\mathbf{x})_i \neq 0} (\mathbf{B}\mathbf{x})_i^2}{\|\mathbf{B}\mathbf{x}\|_0} = \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{B}\mathbf{x}\|_0}$$

Taking square roots and combining this with Lemma 3,

$$\|\mathbf{B}\mathbf{x}\|_{\infty} \geq \frac{\|\mathbf{x}\|_2}{\sqrt{\|\nabla_{\mathcal{T}} \mathbf{x}\|_0 \lceil \log d \rceil \lceil \log n \rceil}}$$

from which we have the result,

$$\|\mathbf{By}\|_\infty \geq \frac{\|\mathbf{x}\|_2}{\sqrt{\|\nabla_{\mathcal{T}\mathbf{x}}\|_0 \lceil \log d \rceil \lceil \log n \rceil}} - \sigma \sqrt{2 \log(1/\delta)}$$

Forcing this lower bound to be greater than  $\tau$  gives us our result.  $\square$

## B Proofs For Section 5

### B.1 Proof of Corollary 11

First we restate Corollary 9 from Von Luxburg et al. [2010]:

**Corollary 15.** *Consider an unweighted symmetric or mutual  $k$ -NN graph built from a sequence  $X_1, \dots, X_n$  drawn i.i.d. from a density  $p$ . Then there exists constants  $c_1, c_2, c_3$  such that with probability at least  $1 - c_1 n \exp(-kc_2)$  we have uniformly for all  $i \neq j$  that:*

$$\left| \frac{k}{2m} H_{ij} - \frac{k}{d_j} \right| \leq c_3 \frac{n^{2/d}}{k^{1+2/d}}$$

*Proof of Corollary 11.* We focus on the symmetric  $k$ -NN graph in which we connect  $v_i$  to  $v_j$  if  $v_i$  is in the  $k$ -nearest neighbors of  $v_j$  or vice versa. In this graph, every node has degree  $\geq k$  which will be crucial in our analysis. Our goal is to bound the effective resistance of every edge, so that we can subsequently bound  $r_{max}$  and apply Corollary 8. From the definition of  $r_e$  we have:

$$\begin{aligned} r_{ij} &= \frac{1}{2} \left( \frac{H_{ij}}{m} + \frac{H_{ji}}{m} \right) \\ &\leq 2c_3 \frac{n^{2/d}}{k^{2+2/d}} + \frac{1}{d_i} + \frac{1}{d_j} \\ &\leq 2c_3 \frac{n^{2/d}}{k^{2+2/d}} + \frac{2}{k} \end{aligned}$$

Where the first line is the definition of  $r_{ij}$ , the second line follows from Corollary 15 and the last line follows from the fact that  $d_i \geq k$  for each vertex. Since  $k(k/n)^{2/d} \rightarrow \infty$ , we see that  $r_{ij} = O(\frac{1}{k})$ . Moreover, with this scaling of  $k$ , that the probability in Corollary 15 is going to 1. We can therefore bound  $r_{max}$  as:

$$r_{max} \leq \rho \left( 2c_3 \frac{n^{2/d}}{k^{2+2/d}} + \frac{2}{k} \right) = O\left(\frac{\rho}{k}\right)$$

Since the first term is going to zero with  $n$ . Plugging in this bound on  $r_{max}$  into Theorem 8 gives the result.  $\square$

### B.2 Proof of Corollary 12

As before, we first state Corollary 8 from Von Luxburg et al. [2010]:

**Corollary 16.** *Consider an unweighted  $\epsilon$ -graph built from the sequence  $X_1, \dots, X_n$  drawn i.i.d. from the density  $p$ . Then there exists constants  $c_1, \dots, c_5 > 0$  such that with probability at least  $1 - c_1 n \exp(-c_2 n \epsilon^d) - c_3 \exp(-c_4 n \epsilon^d) / \epsilon^d$ , we have uniformly for all  $i \neq j$  that:*

$$\left| \frac{n \epsilon^d}{2m} H_{ij} - \frac{n \epsilon^d}{d_j} \right| \leq \frac{c_5}{n \epsilon^{d+1}}$$

*Proof of Corollary 12.* Some manipulation of the result in Corollary 16 reveals that:

$$H_{ij} \leq \frac{2m}{d_j} + \frac{2c_5 m}{n^2 \epsilon^{2d+2}}$$

Under our scaling, the second term goes to zero and the probability in Corollary 16 goes to one, so  $H_{ij} = O(m/d_j)$ . We will now give a lower bound on  $d_j$ . If  $X_i$  is in the ball of radius  $\epsilon$  centered at  $X_j$ , then we connect  $X_i$  and  $X_j$ . Thus  $d_j$  is exactly the number of vertices in the  $B(X_j; \epsilon)$ . The regularity condition on  $p$  in Von Luxburg et al. [2010] requires that there exists constants  $\alpha$  and  $\epsilon_0$  such that for all  $\epsilon < \epsilon_0$  and for all  $x \in \text{supp}(p)$ ,  $\text{vol}(B(x; \epsilon) \cap \text{supp}(p)) \geq \alpha \text{vol}(B(x; \epsilon))$ . By this fact, the fact that the density is lower bounded by  $p_{\min}$ , and by the fact that  $\epsilon \rightarrow 0$ , we know that for sufficiently large  $n$ ,  $p(B(X_j; \epsilon)) \geq p_{\min} \alpha c_d \epsilon^d$  where  $c_d \epsilon^d$  is the volume of a  $d$ -dimensional ball of radius  $\epsilon$ . The probability that  $X_i \in B(X_j; \epsilon)$  is distributed as a Bernoulli random variable with mean  $\geq \alpha p_{\min} c_d \epsilon^d$ . By Hoeffding's inequality and a union bound we get that:

$$d_j \geq n \alpha p_{\min} c_d \epsilon^d + \sqrt{n \log(n)} = \Omega(n \epsilon^d)$$

for all vertices  $j$  with probability  $\geq 1 - 1/n$ . Using the definition of  $r_{i,j}$  along with the bound on  $H_{ij}$  and  $d_j$  we have that uniformly for all pairs  $i, j$ :

$$r_{i,j} = O\left(\frac{1}{n \epsilon^d}\right)$$

Plugging in this bound into Theorem 8 gives us the result. □